Naval Research Laboratory

Washington, DC 20375-5320

Navy Fuel Composition and Screening Tool (FCAST) v2.8

Mark H. Hammond
Robert E. Morris
Jeffrey A. Cramer
Thomas N. Loegel
Kevin J. Johnson

Navy Technology Center for Safety and Survivability
Chemistry Division

Kristina M. Myers

Nova Research, Inc.
Alexandria, Virginia

May 10, 2016

REPORT DOCUMENTATION PAGE

1. REPORT DATE *(DD-MM-YYYY)* 10-05-2016	2. REPORT TYPE Memorandum Report	3. DATES COVERED *(From - To)* 4 April 2011 – 31 January 2016

4. TITLE AND SUBTITLE

Navy Fuel Composition and Screening Tool (FCAST) v2.8

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Mark H. Hammond, Robert E. Morris, Jeffrey A. Cramer, Thomas N. Loegel, Kevin J. Johnson, and Kristina M. Myers*

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
61-9251-B-5-5

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Research Laboratory
4555 Overlook Avenue, SW
Washington, DC 20375-5342

8. PERFORMING ORGANIZATION REPORT NUMBER

NRL/MR/6180--16-9685

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research
One Liberty Center
875 N. Randolph St., Suite 1425
Arlington, Virginia 22203-1995

Naval Air Systems Command
Air-4.4.5
Patuxent River Naval Air Station
Patuxent River, Maryland 20670-1534

10. SPONSOR / MONITOR'S ACRONYM(S)

ONR, NAVAIR

11. SPONSOR / MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

*Nova Research, Inc., 1900 Elkin Street, Suite 230, Alexandria, VA 22308

14. ABSTRACT

The high cost and limited availability of emerging alternative fuels are often major impediments to certification of these fuels as Fit-For-Purpose (FFP) for the U.S. Navy. A method whereby a candidate fuel could be rapidly screened for many FFP properties, using a minimal volume (<1 mL), would overcome this limitation. To meet this challenge, algorithmic modeling strategies were derived that establish the statistical relationships between composition and critical FFP fuel properties. This has allowed us to develop partial least squares (PLS) models based on gas chromatography–mass spectrometry (GC-MS) data that predict fuel properties. The Fuel Composition and Screening Tool (FCAST) was developed as a general fuel screening tool that combines GC-MS based property predictions with a compositional profiler to provide a variety of useful information about a fuel sample.

15. SUBJECT TERMS

Navy mobility fuels	Gas chromatography–mass spectrometry	Partial least squares
Fit-for-Purpose	GC-MS	PLS
Certification	Chemometric property modeling	Compositional profiler

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Mark Hammond
a. REPORT Unclassified Unlimited	**b. ABSTRACT** Unclassified Unlimited	**c. THIS PAGE** Unclassified Unlimited	Unclassified Unlimited	41	**19b. TELEPHONE NUMBER** *(include area code)* (202) 404-3354

CONTENTS

FIGURES

profile in area percent; 8) Chemical structure of the selected compound in the hydrocarbon profile.

16. FCAST Hydrocarbon Distribution screen, showing 1) List of data files; 2) Carbon number distributions in area percentages for different classes of hydrocarbons in the sample; 3) Bar graph depicting the carbon number distributions in a selected compound class (selectable via the upper table).

17. FCAST Hydrocarbon Distribution screen, showing 1) List of data files; 2) Stacked Bar graph depicting the carbon number distributions in the selected compound classes 3) Compound class list as check boxes to add or remove from bar chart.

18. FCAST Distillation Curve screen, showing 1) List of data files; 2) Predicted distillation curve shown in black, along with jet and diesel reference curves.

19. FCAST Label Peaks screen, showing 1) List of data files; 2) TIC with labels based on profile; 3) Selection tree enabling choices of either compound classed, or individual compounds.

TABLES

1. Properties predicted by the FCAST.

ABBREVIATIONS

ASTM   American Society for Testing and Materials

CUMPRESS   Cumulative predicted residual error sum of squares

DiEGME   Diethylene glycol monomethyl ether

FCAST   Fuel Composition and Screening Tool

FFP   Fit for purpose

FSII   Fuel system icing inhibitor

GC   Gas chromatography

GC×GC   Comprehensive two-dimensional gas chromatography

GC-MS   Gas chromatography with mass selective detection

IUPAC   International union of pure and applied chemistry

LOO-CV   Leave one out cross validation

LV   Latent variable

MF   Match factor used in MS library searching

MS   Mass spectrum

MSD   Mass selective detector

NaN   Not a number (null)

NFPM   Navy Fuel Property Monitor

NIST   National Institute of Standards and Technology

NRL   Naval Research Laboratory

PLS   Partial least squares

RMSECV   Root mean squared error of cross validation

SVD   Singular value decomposition

TFA   Target factor analysis

TIC   Total ion chromatograph

UVE-PLS   Uninformed variable elimination partial least squares

XML   Extensible markup language

1.0 Introduction

The high cost and limited availability of emerging alternative fuels is often a major impediment to certification of these fuels as Fit-For-Purpose (FFP) for the U.S. Navy. A method whereby a candidate fuel could be rapidly screened for many FFP properties, using a minimal volume (< 1 mL), would overcome this limitation. The Navy Fuel Property Monitor (NFPM) was a screening tool developed for shipboard quality surveillance, based on chemometric modeling of near-infrared (NIR) spectra. While this is has proven to be a viable approach for known (calibrated) fuels, spectral modeling is not practical when applied to fuels that are radically different in composition (uncalibrated), from those used to derive the models. Thus, spectral modeling was deemed impractical as a tool to model properties of alternative fuels and/or blending stocks with unknown compositions.

In order to meet this challenge, algorithmic modeling strategies were derived that establish the statistical relationships between composition and critical FFP fuel properties. This has allowed us to develop partial least squares (PLS) models based on gas chromatography-mass spectrometry (GC-MS) data that predict fuel properties more accurately than NIR. More significantly, these models are also capable of predicting critical specification properties of blends of Navy mobility fuels with new alternative fuels, regardless of their source or processing methods.

The Fuel Composition and Screening Tool (FCAST) was developed as a general fuel screening tool that combines GC-MS based property predictions with a compositional profiler to provide a variety of useful information about a fuel sample.

This document is an update to the previous NRL Memorandum report[1], which includes additional features incorporated in FCAST version 2.8.

2.0 Fuel Characterization by GC-MS

2.1 NRL Compositional Profiler

The NRL compositional profiler is an automated chemical component classification tool that was developed to provide a classification of all compound classes in a fuel, as an alternative to ASTM D2425[3], which does not function adequately with alternative, non-petroleum derived fuels. The profiler has been implemented in the Navy protocols for alternative jet[4] and diesel[5] fuel certification. The profiler functions by reading the GC-MS data file, identifying each unique compound peak, performing a noise analysis, then sending the peak table to a NIST electron impact mass spectral library. The chemical compounds thus identified are classified with respect

to a set of 25 defined compound classes using a set of selection rules that operate on either the molecular formula or IUPAC name. In addition, the profiler calculates and reports carbon number distributions, average carbon number and degrees of unsaturation for each carbon number.

The accuracy of the NRL compositional profiler has been verified with known fuels and surrogate fuel blends. In addition to certification, the profiler has proven to be a useful tool for rapid interpretation of GC-MS fuel analyses and is being employed in the FCAST program to provide compositional data for the statistical modeling.

The profiler classifies all detectable compounds in the sample with respect to the following compound classes:

Saturates
- Normal Alkanes
- Iso Alkanes
- Monocyclo Alkanes
- Alkyl Monocyclo Alkanes
- Dicyclo Alkanes
- Alkyl Dicyclo Alkanes
- Tricyclo Alkanes
- Alkyl Tricyclo Alkanes

Olefins
- Acyclic Alkenes
- Cyclo Alkenes

Aromatics
- Alkyl Benzenes
- Indans and Tetralins
- Indenes
- Naphthalenes
- Branched Naphthalenes
- Acenaphthenes
- Acenaphthylenes
- Tricycloaromatics

Heteroatomics
- Methyl Esters
- Sulfur-Bound
- Nitrogen-Bound
- Oxygen-Bound
- Chlorine-Bound
- Other Halogen-Bound

Other (not in above classes)

A prefilter is also used to pre-emptively remove known interferents, e.g., polysiloxanes, which are associated with bleed from the GC column stationary phase and methylene chloride, commonly used as a solvent.

The NRL Compositional Profiler has been demonstrated[6,7] to be effective for rapid compositional profiling of complex mixtures that would otherwise take unreasonable times to manually analyze. Nevertheless, a major limitation of the standalone NRL profiler algorithm is that it reports relative contribution of each class of compounds as a percentage of the total area counts measured in each analysis, thus neglecting the effect of differing response factors among different compounds on the GC-MS. Compound-specific response factors observed with GC-MS are dependent on the molecular ionization efficiency of the compound, and to some extent, the fragmentation pattern induced by the mass spectrometer. This makes them highly dependent on molecular structure in ways that are difficult to generalize across a wide range of possible mixture constituents[8].

While peak area abundances are self-consistent within a given sample or group of similar fuels, it is not always possible to mathematically operate with such area based profiler results, when comparing as alternative and petroleum fuels. Response factors were empirically derived by collecting both the FID and MS responses from a GC-FID-MS instrument using standards for each compound class. The FID and MS responses were then compared and then an average for each compound class was used to derive the MS response factor. The normal alkane class was used as a baseline and given a response factor of 1.0. Additional classes tested included iso-alkanes, cyclo-alkanes, olefins and aromatics.

The Profiler saves the data based on the area of the TIC, and adjusts the abundance of each compound, based on carbon number and class. This enables the compound class profiler in the FCAST to report compound abundances in mass percent.

2.2 *Modeling Fuel Properties from GC-MS Data*

It is known[9-12] that a great deal of information regarding fuel composition can be obtained from GC-MS, and the wide availability of this instrumentation make it an ideal analytical technique upon which to base a fuel modeling tool. Compositional information can be derived from the analysis of GC data or GC×GC data without the benefit of mass spectrometry[13-16] and from MS without the benefit of chromatography[17], as well as GC-MS data and GC×GC-MS data without the benefit of complete mass/charge ratio information[18-23]. Nevertheless, fuel-based FFP modeling requires the discrimination of hundreds of discrete compounds, and gas chromatography-mass spectrometry (GC-MS) has the potential to provide this level of discrimination. Algorithms such as Target Factor Analysis (TFA)[24], instrumental modes such as selected ion monitoring[25], or comparative techniques requiring the use of internal standards[26] can

be used to interrogate GC-MS data sets for individual target compounds. However, attempting to explicitly target every compound that could potentially be found in a fuel sample is not realistic. Previous multi-way modeling[27] performed in this laboratory focused on elucidating the compositional differences between different fuel samples. The data features quantified in that work were the same type of data features from which fuel constituencies would be derived in this study, but a more direct focus on fuel constituency is necessary for direct fuel property modeling.

GC-MS data are represented by a 3-dimensional array[28] consisting of (mass/charge) vs (abundance) vs (chromatographic retention time). In order to apply a PLS analysis to this data, it is first necessary to represent it with an appropriate 2-dimensional abstraction. This was accomplished by the construction of an n-dimensional abstraction vector, where *n* equals the number of discrete chemical compounds found in our worldwide fuel calibration set. Each element of this abstraction vector is assigned to a different specific compound and the magnitude of each element represents the abundance of that compound in the fuel sample. This 2-dimensional abstraction vector that represents the fuel composition can be referred to as a metaspectrum of compound vs abundance, as shown graphically in Figure 1. In order to construct the metaspectra, the TIC peaks are identified and, then sent to the NIST Mass Spectral search engine. Compound abundances are calculated from the TIC peak areas and these peak areas used to set the magnitudes of the appropriate elements in the abstraction vector. With appropriate peak thresholds, the vast majority of chromatographic artifacts and masking compounds are automatically eliminated. It was determined that a peak area threshold of 0.001% was the best choice for the most analyses.
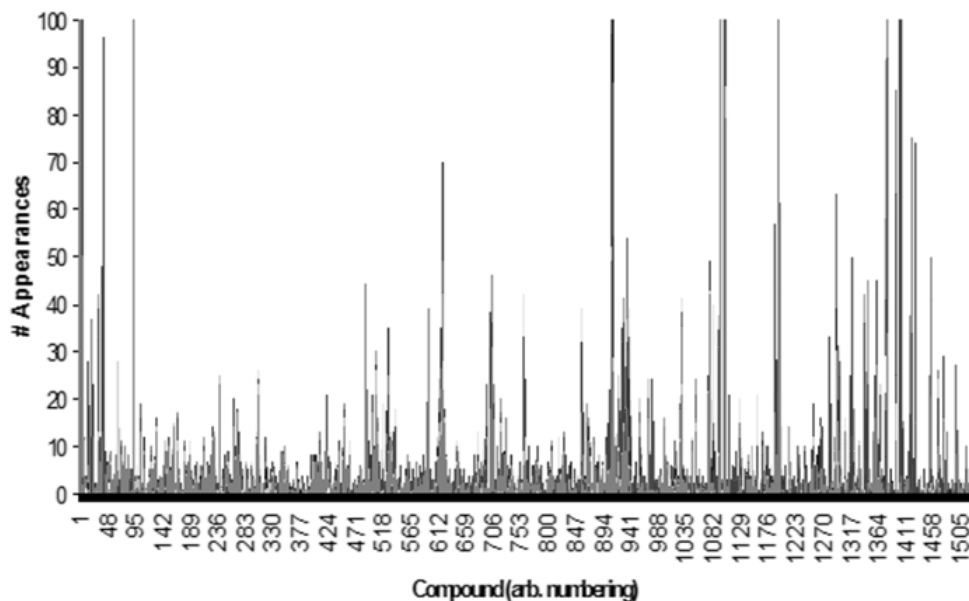

Figure 1. Plot of the abstraction vector, which is a two dimensional metaspectral representation of fuel composition.

Because it would be impossible to predict every possible compound that could be present in a future fuel population, and to produce a training data set that would allow for their future identification, a methodology was developed to more ably accommodate those compounds that do not appear to any significant extent in the training data set. Instead of simply using the best possible NIST database identity match alone, the second best possible identity is also considered. These identities are then each compared to a master list of compounds that were actually found during the production of the original fuel property prediction models. If the first most likely compound does not appear in the master list, then the second most likely compound can be tested. In this fashion, uncalibrated compounds that are nonetheless structurally similar to calibrated compounds won't be ignored and are allowed to influence fuel property prediction results in an appropriate fashion.

PLS Regression Analysis. PLS regression was used to develop the statistical correlations between the component spectra of the fuel samples to their measured ASTM fuel property values. The technique of PLS is based on singular value decomposition (SVD), which mathematically transforms data based on the underlying linear variances that can be found within it. SVD results in a linear transformation of the data into new variables, known as latent variables (LVs) because they are not directly observable in the original data. These LVs are calculated so as to maximize covariance between the data and the variable to be predicted, which allows the differentiation of larger and smaller sources of variance not only from each other, but also from possible interfering factors, producing multivariate prediction models that provide a higher level of overall model robustness than can be afforded by simple univariate prediction models.

It is critical to choose the appropriate number of LVs to use in a particular property model. The trade-off is one of bias versus variance: if too few LVs are used, the model will inadequately model the property of interest, producing biased predictions, while if too many are used, the model will overfit to spurious variance in the calibration data and poorly predict the properties of new, uncalibrated sample data. Achieving this balance between modeling precision and robustness is particularly challenging when modeling fuel properties, due to the variable nature of fuel compositions. The number of LVs to be retained in each PLS fuel property model construction were determined using leave-one-out cross-validation (LOO-CV)[29] which approximates model performance with uncalibrated data. In LOO-CV, the predicted fuel property value of each fuel sample in a given model is based on a sub-model built from every other sample except for the sample being given a prediction value. This operation produces a single Root Mean Square Error of Cross-Validation (RMSECV) result for each number of LVs evaluated. Choosing the number of LVs that minimizes this RMSECV value theoretically maximizes the performance of a given model with uncalibrated data. However, RMSECV results are ultimately an imperfect metric to use to optimize the number of LVs in this type of modeling.[20-32] This is because RMSECV results are still based on models that take almost all of the available training data into account and are, therefore, being created under the assumption

that the training data are completely representative of all possible future data. This assumption may be valid when only modeling petrochemical fuels, but is not necessarily valid with the inclusion of alternative fuels of unknown compositions in sample populations.

To compensate for this reality, the numbers of LVs to be used for each fuel property prediction model produced were instead chosen automatically using an F-test statistic.[33-35] The F-test was applied to the LOO-CV cumulative predicted residual error sum of squares (CUMPRESS) results with an 85% confidence interval, using a maximum of 10 LVs. The use of the F-test tends to select a smaller number of LVs than the minimum RMSECV value would suggest, which, in turn, sacrifices the immediate quality of a model in order to better preserve its robustness and its potential utility in the presence of uncalibrated data. By limiting the number of LVs that can be incorporated into a model, the F-test protects against overfitting. Once the number of LVs was chosen using the F-test, each model was rebuilt using all possible calibration data to obtain the final Root Mean Square Error of Prediction (RMSEP) results.

Uninformative Variable Elimination PLS. A modified version of PLS known as UVE-PLS[36] was used in this work to remove variables from the PLS model training data that contribute minimal or no relevant information toward the given modeling goal. With GC-MS derived metaspectra, this results in the elimination of uninformative individual compounds, focusing the construction of the PLS models on those compounds that are most statistically significant with respect to the fuel property being modeled. Although the elimination of specific compounds may seem counterintuitive, given the stated goal of developing a comprehensive FFP modeling strategy, the eliminated compounds deemed to be uninformative through UVE-PLS were not contributing constructively to model quality, and thus constituted only noise or interference. Furthermore, because the models produced through the use of UVE-PLS still retained many compounds, regardless of the strategy being used, it is our hypothesis that these models will still be capable of accommodating future fuels, regardless of their composition.
In order to understand how UVE-PLS functions, first consider the basic equation for Partial Least Squares:

$$\mathbf{y} = (\mathbf{X_c}*\mathbf{b}) + \mathbf{e} \tag{1}$$

where, $\mathbf{y}$ is the (n x 1) vector of calibration values (in our case, fuel properties), one for each of the n fuel samples; $\mathbf{X_c}$ is the (n x p) data used to predict the calibration values (in this case, our metaspectral data, one vector of length $\mathbf{p}$ per sample because there are $\mathbf{p}$ possible compounds); $\mathbf{b}$ is the (1 x p) vector of regression coefficients that is obtained by using PLS; and $\mathbf{e}$ is the (n x 1) error vector (i.e. the data variance not described by the regression coefficients).

The actual (n x 1) vector of fuel property predictions one obtains from the PLS model ($\hat{\mathbf{y}}$) can be summarized as:

$$\hat{\mathbf{y}} = (\mathbf{X_n}*\mathbf{b}) \tag{2}$$

where, $\mathbf{X_n}$ is the data for the new sample to be analyzed, and $\hat{\mathbf{y}}$ is a vector of fuel property predictions. UVE-PLS requires a cross-validation procedure. In this work, as described previously, a leave-one-out cross-validation was used. Each step in the leave-one-out cross-validation produces not only a separate $\hat{\mathbf{y}}$ vector of fuel property predictions (from which to ultimately calculate RMSECV values), but also a separate $\mathbf{b}$ vector of regression coefficients. Each of these vectors is the length of the number of compounds considered during the modeling procedure, which means that each compound is associated with a set of regression coefficients, consisting of one regression coefficient obtained from each step in the cross-validation procedure. Since each compound has a separate set of regression coefficients, they can then be averaged and assigned a standard deviation value. The ratio of the average over the standard deviation is defined as the *reliability ratio* of that particular compound in the context of a particular PLS property model.

If random variables (compounds) are then added to the list of compounds for a given fuel sample (the original $\mathbf{X_n}$ data set), and thus added to the abstraction vector, then the reliability ratio described above can be calculated for them as well. This is done by adding a number of random variables to the abstraction vector equal to one-fifth of the number of compounds found in that fuel. By comparing the reliability ratios of the added random variables with the actual compounds found in the fuel, it is possible to determine if a given fuel constituent is more informative to a particular model than a random compound.

In this manner, each compound detected in a fuel is tested to determine if its reliability ratio is higher than at least 85% of the random-variable reliability ratios. If it is equal to or greater than 85%, then that compound is retained in the final model. Otherwise, it is removed, since it is inconsistently informative and not contributing to that particular property model.

2.3 Comparison Functions

Two gas chromatography – mass spectrometry (GC-MS) data comparison strategies recently developed and implemented in the Fuel Composition and Screening Tool (FCAST) software produced by the Naval Research Laboratory (NRL). The deltaCompare sub-routine was designed to quickly and quantitatively compare two fuel samples, while the feature selection strategy based on Analysis of Variance (ANOVA) was designed to use the relative differences between larger replicate data populations to isolate more subtle yet still informative data features for further analysis and assessment. It is shown that both comparison strategies produce different but complementary sets of results, and that both sub-routines can find uses in many aspects of fuel analysis.

deltaCompare. This novel computational strategy was developed to provide quantitative information regarding compositional differences of all detected compounds in two fuels. It is a simplified GC-MS comparison strategy that only considers the area-normalized total ion

chromatograms (TICs) of the two fuel samples to be compared. At each individual retention time, the magnitude of the TIC for the two fuel samples are compared, and if the difference between the two values is greater than the standard deviation of the differences in the two TICs at all retention times multiplied by a constant value, then the higher-magnitude mass spectrum corresponding to that retention time is subjected to a NIST database search for identification purposes. This is roughly equivalent to operating on peak differences that are statistically significant with respect to the overall signal to noise ratio of the data. In order to minimize false identifications, it is generally recommended that the constant value be set at 2.33, which is consistent with a one-tailed statistical z-test at a conservative 99% confidence interval (CI). However, there are provisions for the user to specify smaller values, at the risk of false identifications, which can however, be of some use in cases where more subtle compositional changes are sought.

ANOVA Fisher-Ratio Feature Selection. A pointwise ANOVA-based feature selection of GC-MS data was also implemented into the FCAST software to elucidate compositional changes between replicate populations of two samples. As implied previously, this is an improved and streamlined implementation of the methodology used in previous modeling studies to uncover the sometimes subtle compositional changes undergone by fuels during thermal stress. The primary difference between the previous work and the present implementation is that the ANOVA feature-selected data subset representing the compositional differences, can be interpreted by the tools in the FCAST instead of the moving-window parallel factor analysis (PARAFAC) as previously used.

This approach is based on comparing the variance between the two sample populations to the variance present within each population in accordance with Equations 3 and 4. The ratio between these two sources of variance corresponds to the well-known ANOVA F-test statistic, also known as a Fisher ratio, or f-ratio. In summary, the ANOVA feature selection algorithm calculates between and within-sample variance estimates at each point in the GC-MS chromatogram, and then uses these to calculate the f-ratio for every data point in the GC-MS data cube.

$$f\text{-ratio} = \frac{\text{between-class variance}}{\text{within class variance}} = \left[\frac{\sigma^2_{class}}{\sigma^2_{err}} \right] \tag{3}$$

$$\sigma^2_{class} = \frac{\sum (\bar{x}_i - \bar{x})^2 \, n_i}{(k-1)} \qquad \sigma^2_{err} = \frac{\left[\sum \sum (x_{ij} - \bar{x})^2 \right] - \left[\sum (\bar{x}_i - \bar{x})^2 \, n_i \right]}{(N-k)} \tag{4}$$

where, k = #classes (# samples); $\bar{x}_i$ = mean of ith class (sample); x_{ij} = i^{th} measurement of class j; n_i = #measurements in i^{th} class; N = #GC-MS spectra

In its most basic form, an ANOVA F-test is used to assess whether or not the means of two or more sample populations are different. The null hypothesis for this test is that the means are the same. The f-ratio is calculated and compared against a critical value derived from an F distribution with the appropriate degrees of freedom and associated with a given significance level. If the f-ratio is larger than the critical value, then the null hypothesis is rejected and it is concluded that the sample means are different within the confidence interval that is the complement of the significance[1] level used in the test (i.e., a significance level of 0.05 leads to a confidence interval of 95%). Thus, larger f-ratios imply a greater certainty that the sample means are different, although, strictly speaking, the significance level of a statistical test is chosen prior to the test and not driven by the data. At present, the FCAST software allows for the manual selection of f-ratios using a slider, which, in turn, allows end-users to customize the certainty associated with ANOVA-based comparison results.

Outside of a purely statistical context, the f-ratio can also be viewed as a heuristic measure of how discernable two sample populations are, and is conceptually similar to other quantitative measures, such as signal-to-noise and chromatographic peak resolution. In this implementation, there are three main factors that influence the ANOVA f-ratio feature selection results: 1) the magnitudes of the difference in chemical composition, 2) the measurement error, and 3) the number of replicates. The magnitude of the difference in chemical composition between two samples at a given location in the GC-MS chromatogram is reflected in the numerator of the f-ratio, while the denominator is essentially an embodiment of the measurement error itself (Eq. 5). Thus, a sample composition difference of a given magnitude measured by an instrument with a given measurement error can be viewed as a signal-to-noise proposition with an implied tradeoff between the two quantities. In other words, either reducing the measurement error or increasing the magnitude of the chemical composition difference would result in a larger f-ratio while a commensurate increase and reduction in one and the other would maintain a constant f-ratio.

$$f\text{-ratio (GC-MS)} \quad = \quad \frac{\text{variance of a peak between different samples}}{\text{variance of a peak between replicates of single sample}} \quad \textbf{(5)}$$

The number of sample replicates influences the ANOVA f-ratio feature selection by influencing the accuracy with which the component variances of the f-ratio are estimated. As the numbers of replicates are increased, the accuracy of these estimates also increases, conferring increased statistical power[2] to the F-test implied by the f-ratio calculation. This means that for a given confidence interval, more replicates will enable chemical differences with smaller "signal-to-noise" to be detected. This is illustrated in Figure 2, where the base 10 logarithm of the f-ratio

---

[1] Significance in this context is defined as the probability of incorrectly rejecting the null hypothesis.
[2] Statistical power is defined as the probability of correctly detecting a real difference between samples.

is plotted against the logit[3] of the p-value. As shown in Figure 2, assuming a 99.9% confidence interval, an F-test with three replicates of each sample requires an f-ratio of at least 74 to detect a sample difference, while one with 10 replicates of each sample requires an f-ratio of only 15. This represents a five-fold reduction in signal to noise requirements for detection without altering the magnitude of the chemical difference or measurement error of the instrument. At least five replicate analyses of each sample are thus generally recommended in order to help ensure that the component variance estimates are reasonable.
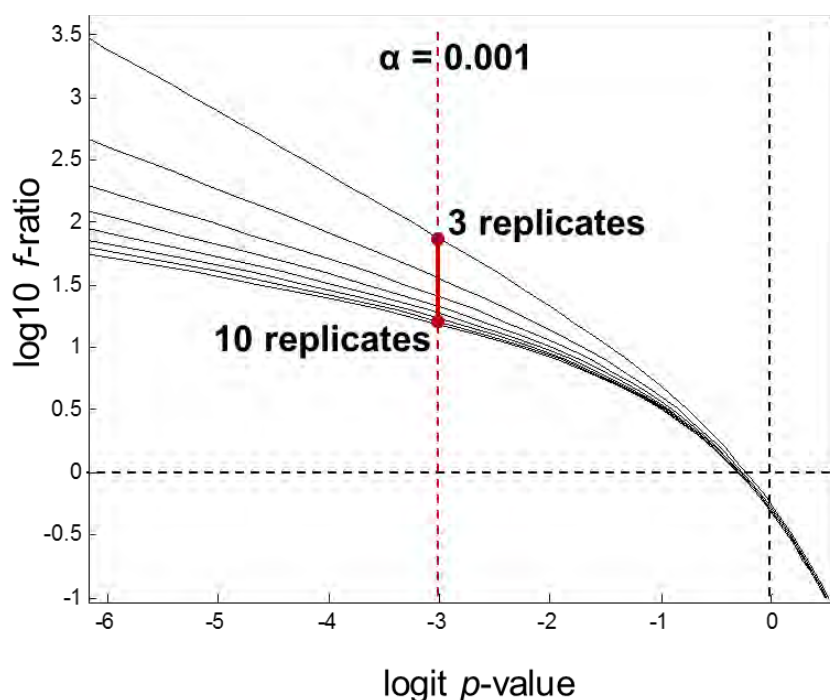

Figure 2. Dependence of the probability of rejecting the null hypothesis on f-ratio.

As was also the case with the deltaCompare sub-routine, this ANOVA-based feature selection approach implicitly assumes that any systematic differences between the two replicate populations are purely due to actual differences in chemical composition. Therefore, practitioners should be careful regarding the potential to introduce non-chemically related systematic differences between the replicate sets; for example, by using widely different GC-MS instruments, or methods to generate the two replicate sets in isolation from each other.

---

[3]Logit is defined as the logarithm of the odds ratio for a given probability, $logit(p) = \log(p/(1-p))$. Logit values less than zero correspond to probabilities less than 0.5 and those greater than zero to probabilities greater than 0.5. Thus, integer intervals on a logit scale represent order of magnitude differences in probability, e.g. the interval (0.001, 0.01) is approximately (-3, -2) on the logit scale.

3.0   The FCAST Software Overview

The FCAST combines the functionality of an improved version of the NRL compositional profiler[1] with modeling of critical fuel properties. The software requires the NIST Mass Spectral Search program to identify the compounds in the sample, and it will notify the user if the spectral database is not installed on the computer. If it is not installed, the software will still run and load Agilent Chemstation GC-MS data files, but no new analyses can be run. The FCAST saves all results of the analyzed data and can display processed results without the original data files. In this way, the FCAST will always show any files that have been analyzed without the need for directory navigation. However, in order to get the results in a readable format, the data must be exported. Analyzed results can be imported from or exported to another computer running the FCAST software.

The FCAST performs all necessary preprocessing and processing steps automatically, and presents the user with the predicted properties, the composition organized by compound class, a lists of compounds found in a fuel, the fuel-relevant compound classes within which these compounds can be classified, carbon number distributions, and other analysis results that may be of interest to various expert and non-expert users. The overall process is illustrated in the flowchart shown in Figure 2.
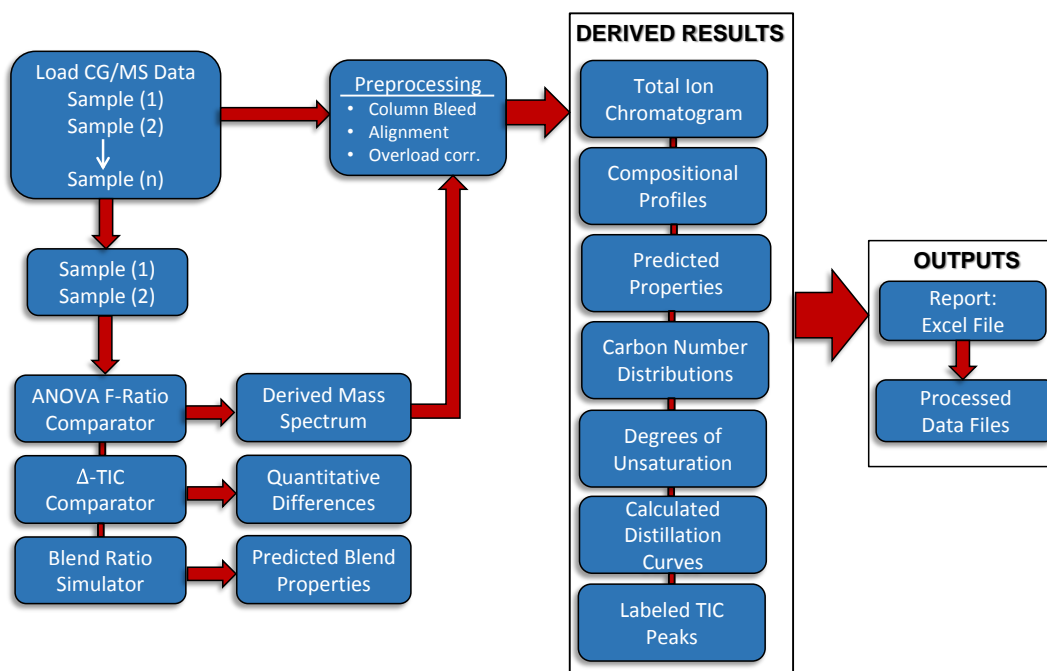

Figure 3. Computational flowchart for the FCAST application.

In the current version of the FCAST software (version 2.6), the compositional profiling is performed independently of the fuel property modeling. This allows a level of versatility in the compositional profiling that is not available in the property modeling because the property

modeling must be performed using a certain set of analysis parameters. Changing, for example, the peak area threshold value to be used with an incoming data set creates a metaspectrum that is fundamentally ill-suited for use in models constructed using different peak area threshold values. However, the compositional profiler, not being similarly restricted, can be used with different minimum peak area thresholds, MF thresholds, and MS data ranges to parse and explore fuel composition in a more free-form manner. The default settings for these parameters are nearly the same as those used for the property modeling, so leaving them unchanged will still result in an effective analysis. A solvent delay can also be an input into the compositional profiler in order to exclude non-compositional, early-spectrum data artifacts.

The property predictions are checked against the models to determine if the data falls within the model to make a valid prediction. If the sample falls outside of the property model the value is reported as NaN (Not a Number), so as to not report a false value. In addition, since the property models were designed for conventional fuels and fuel blends containing alternative fuels, these models are not accurate when applied to compositionally sparse or pure compounds. Thus, if any single component makes up more than 80% of the sample no property calculations are performed. The fuel properties that are predicted by PLS modeling of the GC-MS metaspectra in the FCAST are shown in Table 1.

Detection of sample overloading. In a GC-MS analysis, there are two ways that a sample can overload the system that will impact the integrity of the final results: 1) Chromatographic overloading, and 2) detector overloading. Chromatographic overloading occurs when the quantity of an analyte on a GC column exceeds the capacity of the stationary phase, causing the analyte to elute in a non-Gaussian manner characterized by non-symmetrical peak shapes and can adversely affect peak area calculations. However, even when column overloading is not evident, GC-MS detector overloading can still occur. This is a consequence of a limitation in the Agilent GC-MS data file format, where the number of ion counts for a particular m/z fragment peak is limited to a maximum value of 8388608. Since the TIC is calculated as the sums of detector counts at each GC retention time, if this variable limit is exceeded, the peak areas will not be correct.

However, software variable overloading of this nature is not always evident from TIC peak shapes and must be explicitly checked. The FCAST software checks each incoming GC-MS data file for variable overloading by checking all m/z values to ensure that they are less than the maximum value (8388608). Any m/z values that are at maximum are classified as "overloaded" and the total percentage of overloaded peaks are calculated as a percentage of the total number of peaks. Overloaded peaks will introduce errors in the relative abundances of the different compounds calculated by the profiler, as well as cause errors in the calculated properties. If the user processes a file that contains overloaded peaks, this will be shown in red in the information box, to the right of the TIC display. It is recommended that overloaded GC-MS data be

reacquired with a lower sample concentration. Functionality to correct mildly overloaded peaks has been added, which looks at the last scan that is not overloaded and uses the ratio of the non-overloaded masses to determine the intensity of the overloaded masses. This approach, however, has limitations if too many masses are overloaded, or if the signal the too overloaded, then an accurate correction will not be achieved.

Direct Calculations. Those fuel characteristics that can be directly calculated from the GC-MS data, are not modeled. Fuel system icing inhibitor (FSII) is required to be added to military jet fuels. FSII, more specifically diethylene glycol monomethyl ether (DiEGME), is a single compound which can be identified by its two most abundant mass fragment ions at m/z=45 and 59. There are a limited number of other typical fuel constituents that produce the same m/z=45 ion, but since those interfering compounds do not also produce the m/z=59 ion, they can be eliminated. Examining these two fragment ions, the software first determines if there is any DiEGME at all and returns zero (0) as a value if not. If DiEGME is determined to be present, then the sum of the two fragment ions are compared to the whole sample and used to determine FSII composition in the fuel.

Distillation curves acquired in accordance with ASTM D86[37] can be modeled from composition, but direct calculations are more precise. A method was developed by which a simulated distillation (SIMDIS) type of calculation can be performed on the GC-MS data without calibration standards. The SIMDIS determines the temperatures at which 10%, 20%, 50%, and 90% of the fuel would be distilled from the fuel. This is accomplished by using the GC retention time indices of identified straight chain alkanes (with known boiling points) as an internal standard. The alkane retention time indices are used to map boiling point to retention time and the distillation points are then calculated as percentages of total fuel eluted from the column. If there are too few identified alkanes to adequately define the boiling point map (e.g., a pure compound or a mixture of two pure compounds), the software will not display the distillation point numbers.

Table 1. Properties predicted by the FCAST.

PROPERTY	Units	ASTM	#Samples	LVs	RMSEP
Collegative					
Density (g/cm3)	g/cm3	D4052	749	9	0.0034
Flash Point	°C	D93	796	7	5.1428
Pour Point	°C	D5949	191	6	5.6025
Freeze Point	°C	D5972	402	9	1.9876
Cloud Point	°C	D2500	351	6	2.5880
Viscosity -20C	cSt	D445	76	2	0.7721
Viscosity 40C	cSt	D445	570	8	0.9007
Acid Number	mg/g KOH	D3242	280	5	0.0296
Cetane Index	--	D976	508	5	1.5487
Dist. IBP (°C)	°C	D86	(a)	(a)	(a)
Dist. 10% (°C)	°C	D86	(a)	(a)	(a)
Dist. 20% (°C)	°C	D86	(a)	(a)	(a)
Dist. 50% (°C)	°C	D86	(a)	(a)	(a)
Dist. 90% (°C)	°C	D86	(a)	(a)	(a)
Dist. FBP (°C)	°C	D86	(a)	(a)	(a)
Constituents-Major					
Olefins	vol%	D1319	61	6	0.2922
Saturates	vol%	D1319	61	7	0.3130
Aromatics	vol%	D6379	84	5	1.8466
Naphthalenes	vol%	D1840	47	4	0.1432
Constituents-Minor					
FSII (DiEGME) (b)	wt%	D5006	(a)	(a)	(a)
Hydrogen	wt%	D3701	83	7	0.4325
Sulfur	wt%	D4294	559	8	0.0788
Karl-Fischer Water	ppm	D6304	50	6	3.8428
Insolubles					
Existent Gum	mg/100 mL	D381	233	3	1.5600
Lubricity (BOCLE)	WSD mm	D5001	253	6	0.0393
Storage Stability	mg/100 mL	D5304	389	7	0.7067
Demulsification	minutes	D1401	405	3	2.9141

(a) Predicted by direct calculation.

(b) FSII calibration is specific to DiEGME; this tool will not detect other icing inhibitors.

4.0   Using the FCAST Software

The FCAST was developed to operate with Agilent GC-MS Chemstation data files. It is important to bear in mind that the information presented by the FCAST is based on library pattern matching and chemometric analyses of the GC-MS data. Thus, the quality of the results obtained will be directly related to the quality of the raw GC-MS data and care must be taken to ensure that the instrument is configured properly and the chromatography and mass detection are functioning properly.

4.1 GC-MS Data Acquisition

The Instrument must be configured as follows:

- **Instrument**: Agilent 7890A GC connected to an Agilent 5975C MSD with a heated transfer capillary line (250 °C)

- **Column:** 60m x .25mm x 0.5 μm Agilent DB-1ms fused silica with a helium flow of 2.0 mL/min.

- **MS Parameters:** Source temperature 250 °C, Quad temperature 150 °C, Scan Mode scanning from 35 – 400 m/z with a threshold of 250 and a gain factor of 1.5.

- **Oven Program:** 40 °C for 2 min, 5 °C/min to 165 °C, 2.5 °C/min to 265 °C, 10 °C/min to 295 °C for 0 min, Total Run Time of 70 min.

- **GC Inlet:** Split mode, 35:1 split ratio, 285°C

- **Sample Preparation:** dilute 5:1 with methylene chloride

- **Injection Volume:** 0.5 μL

It is <u>imperative</u> that the GC-MS method used when generating data for the FCAST be identical or as similar as possible to that used to collect the training data upon which the models and direct property calculations are based. Slight deviations between GC-MS instruments will exert minimal impact on the accuracy of the compositional profiler, as the peak-based mass spectral abstraction process is relatively robust with respect to calibration transfer between different similar instruments. However, the precision of the property models is sensitive to the acquisition parameters used, since they are based on compositional distribution data.

4.2 Data Processing System Requirements

The FCAST application is designed to run on the Microsoft Windows platform and the .NET 4.0 Framework. It will run on both 32- and 64-bit versions of Windows. A minimum screen resolution of 1024×768 pixels is required and the FCAST software also requires that the NIST Mass Spectral Search Program (NIST11) be installed in the default directory (C:\NIST11\MSSEARCH). If the NIST11 program is not installed, the FCAST software will still run and load samples (even those imported from other systems) but no new analyses can be conducted. When data are being processed using FCAST, any running instance of NIST11 will be terminated, so that a clean start can be used for the analysis. Additionally NIST11 will terminate when the analysis is complete and the NIST11 history files will be cleared.

4.3 Software Installation

The FCAST Installation program will install the FCAST program and check whether the .NET 4.0 Framework is installed. The Installer will NOT check if NIST11 is installed as the software does not need it to run, but is necessary to process data.

If you are upgrading from a previous version of FCAST, the Installer will remove the previous version. Any data files created by the program will remain, and the preferences will be maintained.

4.4 Interface Design

The interface is designed to be user friendly with the ability to open data folders and select which files to process. The program also displays previously processed data, eliminating the need to reprocess the data to display results. The interface displays many of the results generated within the program. In addition to the predicted fuel properties, the complete output of the compositional profiler is displayed in a tree format that allows the operator to expand the tree to display the individual compounds detected with estimated abundances in normalized volume percent. Once a compound is selected, the eluted peak is indicated by a marker on the plot of the total ion chromatogram (TIC), as well as the mass spectrum, chemical structure, and NIST library match factor. A tabbed interface also allows the user to view the carbon distributions of the total fuel, as well as in the aromatic, saturated and olefinic fractions. Degrees of unsaturation are also shown, organized by carbon number. Additional tabs show the carbon distribution by hydrocarbon class, an approximated distillation curve, and a TIC display with user selectable compound labels.

The program uses a simple method to identify the peaks in the TIC. For samples such as hydrocarbon fuels, the compounds involved are too similar to separate by considering the mass

fragment information. The identified peaks are sent to the NIST database for identification, resulting in computation times on the order of 5 minutes per sample for typical petroleum based fuels. To account for uncalibrated compounds not in the master list for property prediction, the second result returned from the NIST search is used. The FCAST also incorporates a data export functionality that allows the operator to select which processed data will be sent to a formatted MS Excel spreadsheet. There is also the option to export the compositional information, including compound classes, carbon numbers and other data to text files for import into other software applications.

All processed data are listed in an XML index and stored in a separate binary database, so that it is not necessary to retain the original raw GC-MS data files to view or reprocess the analysis results. It is only necessary to retain the raw data if it is desired to maintain the original context for the data. When the user selects a directory containing Agilent Chemstation data, or selects the processed data folder, the names of the files in that directory are shown in the pane on the left. Selecting any of the file names immediately displays the TIC and file header information. The user has the option of processing selected files or all files and once processed, they file names are displayed in bold text

5.0   Menu Commands

5.1 File

Load Data*:* Select the folder containing the Agilent Chemstation GC-MS ".D" files to process. While loading the list of files in the directory, hitting ESC will stop at the files loaded.

Recent Folders*:* Keeps track of the last 5 folders selected

Export Results: Saves a comprehensive summary results from all processed samples in an Excel spreadsheet "Summary-MM-DD-YYYY hh-mm-ss.xlsx" in the folder containing the ".D" data files. The spreadsheet contains a summary tab with the calculated property results and compositional profiler results for all samples, and a separate tab for each fuel sample that contains the above results for that sample, in addition to the entire list of identified compounds from the NIST search. Once the export begins, the operation can be canceled, but this will not export any results.

The first window (Figure 3) allows for selection files to export. Once the export is complete, the user has the option to open the file in Excel, or return to the FCAST program.
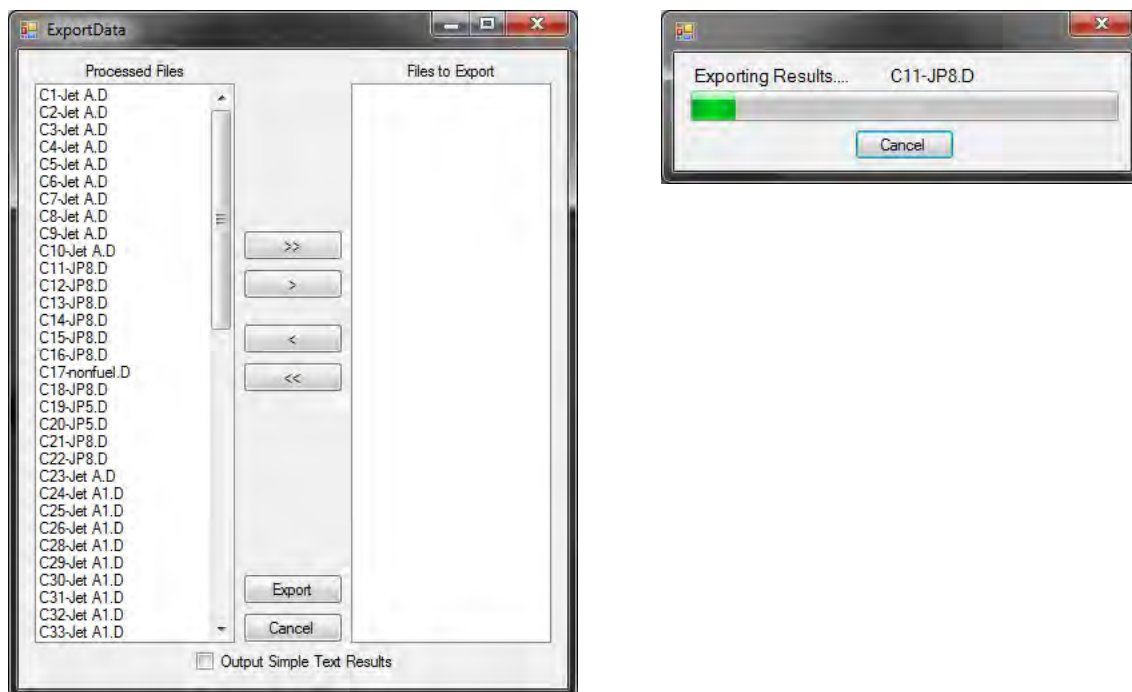
Figure 4. File export dialog window and export progress bar.

Processed Files: This menu item allows for importing, exporting and viewing processed data files.

- **Import***:* Imports the XML and processed data files previously exported from another computer running the FCAST software. The software will ask for a folder location containing the file and copy them into the user directory.

- **Export***:* Exports the XML and processed data files to a folder, for moving to another computer running the FCAST software. The user will be prompted for which files to export from the list of ALL processed files (not just the directory selected). Then the user will be prompted for which folder (or create a new folder) to save the files.

- **View:** This option will populate the list of files stored in the database that are already processed (but without knowing the original location). All other options of seeing the GCMS data, Properties, and Profile are available, as well as reprocessing the file.

- **Delete Entries:** This well remove processed data files from the database, this will not remove the original data files located outside of the FCAST software.

Exit: Exits the FCAST program. The user will first be prompted to confirm.

5.2 Process

The FCAST software analyzes the TIC for identifiable peaks meeting the minimum area criteria. These peaks are then sent to the NIST MS Search program for identification. The returned results are then screened, selecting those that are above the minimum match factor criteria. All the results are saved so that the minimum match factor can be changed without needing to reprocess the file. These compounds are then sent to the profiler to determine which class they fall into. Additionally, the results of the NIST search are used to calculate the properties of the fuel. The status of the analysis is shown at the bottom of the window indicating how many peaks are being analyzed and approximately how long the current processing should take.

Selected File(s): Process the currently selected file(s)

New Files: Process all unprocessed files

All Files: Process all files in the current directory

Cancel Processing: Terminate the analysis, without saving any information. This will also terminate the current instance of the NIST MS Search Program.

Reduced Profile: Once the fuel is profiled, the user can select a section of the sample, based on retention time, and view the compounds in only that section of the sample. The percentages listed will still be based on the entire sample and not only from the reduced set of compounds. This feature is useful to selectively examine, for example, heavy contaminants in a fuel sample.

Blend Fuels: Experimental feature, allows for calculating properties of mixes of two fuels (at 10% intervals). This option is available when the processed files are displayed, since the user first chooses two fuels to test blending (Figure 4). Once processed a new window (Figure 5) shows TIC for fuel A, fuel B and blended fuel, a slider to choose blend level. A table of properties for each step of the blended fuel is also shown.
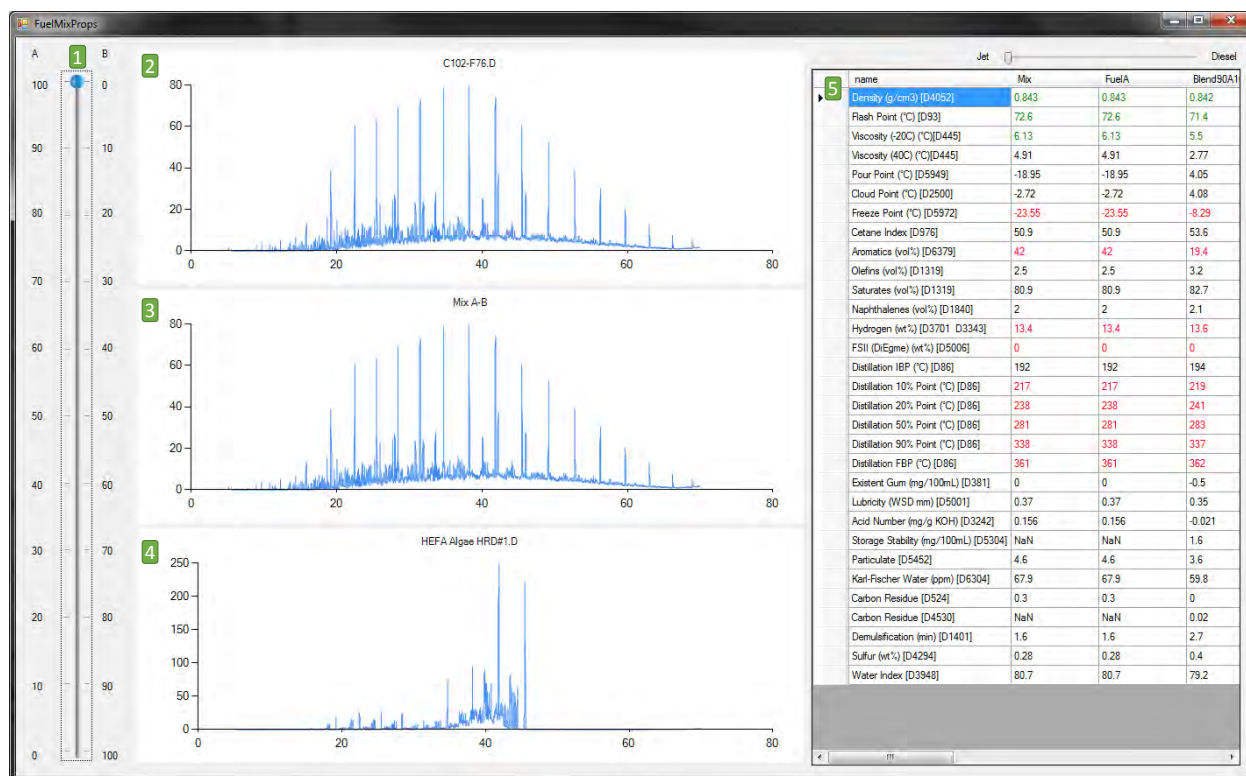

Figure 5. Blend Fuels selection window.

Figure 6. Blended Fuels results window, (1) allowing mixing percentage, (2) Fuel A, (3) Mixed Fuel, (4) Fuel B, (5) Blended properties.

ANOVA: This function allows the comparison of two fuels using the Fisher Ratio[38], which is derived from an analysis of variance (ANOVA). For the ANOVA procedure, a minimum of 5 replicate GC-MS analyses are required for each fuel sample to be compared. The data files available are those in the directory selected by the main FCAST window. Figure 6 shows a screenshot of the ANOVA comparison tool. The left listbox (1) shows all the files available to compare. The two list boxes middle are the samples selected as class A (2) and B (3) for comparison. The two buttons labeled '>' add files to each class respectively, and the buttons labeled '<' remove samples. The button 'A <-> B' swaps the samples used for each class, which is useful if the alignment step does not give good results. Since the ANOVA operates on all data points, proper alignment of the replicate spectra is critical in order to avoid errors. In the ANOVA subroutine, all the samples are aligned to the first sample in class A, which in some cases, can result in misalignment.

The checkboxes for Normalize and Align allow the user to select whether or not those options are enabled. The Analyze button will load the data, then normalize and/or align, if selected, then show a plot of the processed data for observation (Figure 7) to allow the operator to ensure that the peaks are properly aligned. If the data are not aligned, the operator has the option of reversing the two classes, or not aligning the data, to obtain proper alignment of the peaks. The feature selected mass spectrum derived from the ANOVA is displayed for the chosen f-ratio (4) which can be adjusted using the slider control (6). The sum of the f-ratios at each retention time is displayed (5), showing where the largest variance between the two samples is located in the spectrum.
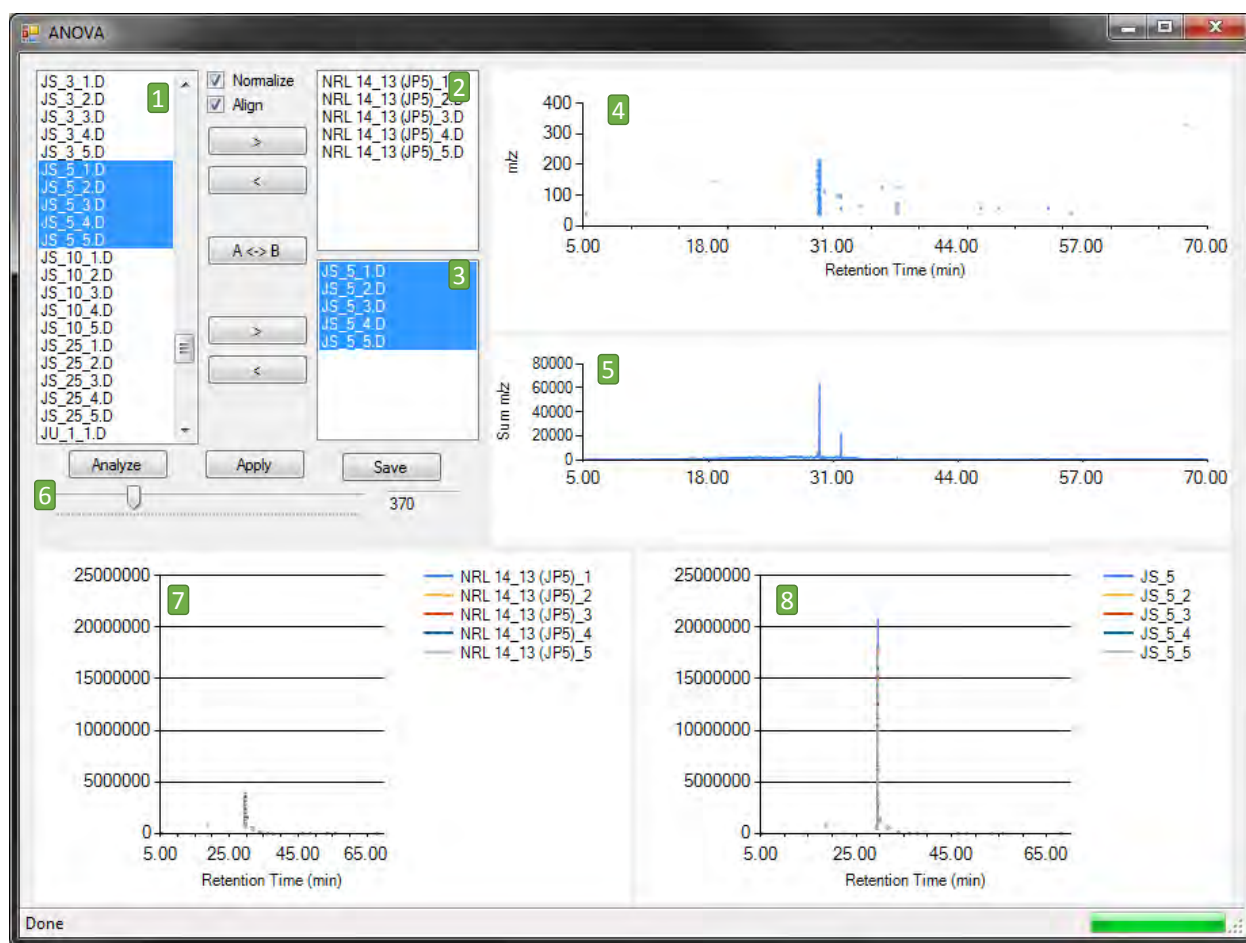
20

Figure 7. FCAST ANOVA screen, showing 1) List of data files; 2) Selected samples for class A; 3) Selected samples for class B; 4) Feature selected mass spectrum based on the selected f-ratio; 5) Sum of the f-ratios at each retention time; 6) f-ratio selector; 7) Feature selected TIC for class A and 8) Feature selected TIC for class B.
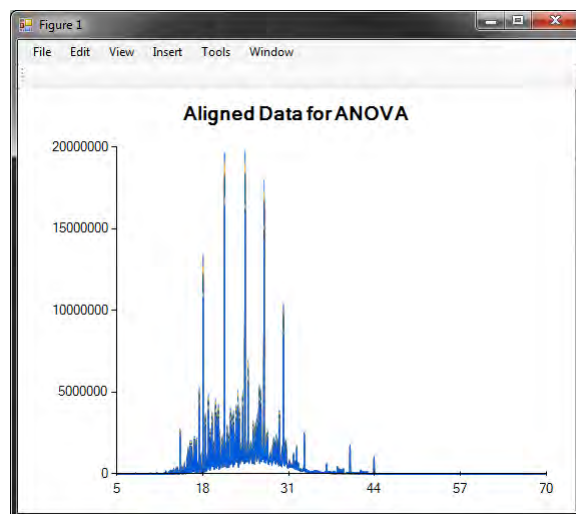

Figure 8. FCAST ANOVA chromatogram alignment window.

21

deltaCompare: The deltaCompare is a simplified GC-MS comparison strategy that only considers the area-normalized TICs of two fuel samples to be compared. The advantage of the deltaCompare is that it only requires one GC-MS analysis per sample. The disadvantage of the deltaCompare with respect to the ANOVA is that instrumental variations are not taken into account. At each individual retention time, the TIC values for the two fuel samples are considered, and if the difference between the two values is greater than the standard deviation of the TIC-based differences at all retention times multiplied by a constant value (two standard deviations), then the mass spectra corresponding to that retention time is reported.
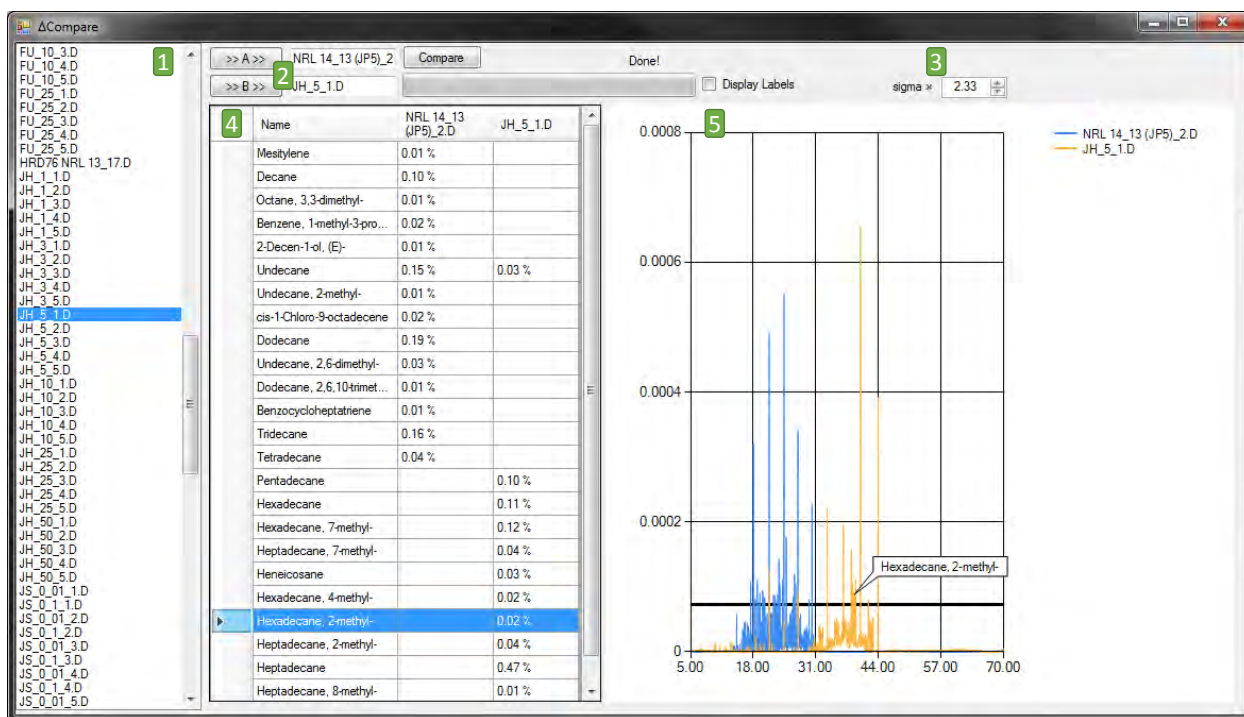

Figure 9. deltaCompare screen, showing 1) List of data files; 2) Selected sample for class A and B; 3) Selected sigma multiplier; 4) Feature selected mass spectrum based on the selected f-ratio; 5) Graph of the A-B and B-A TIC showing identified components.

Dendrogram: The dendrogram function provides a means for the operator to compare a set of replicate GC-MS data, or GC-MS data from different fuel samples. The dendrogram is a simplified hierarchical analysis based on the first two principal components from a PCA cluster analysis of the submitted GC-MS data. The distance between two dendrograms on the x-axis is indicative of the similarity in the samples and can be used to determine if a set of replicates of samples are suitable for use in the comparison functions available in FCAST. It can also serve as a means to classify a set of fuel samples with respect to their compositions.
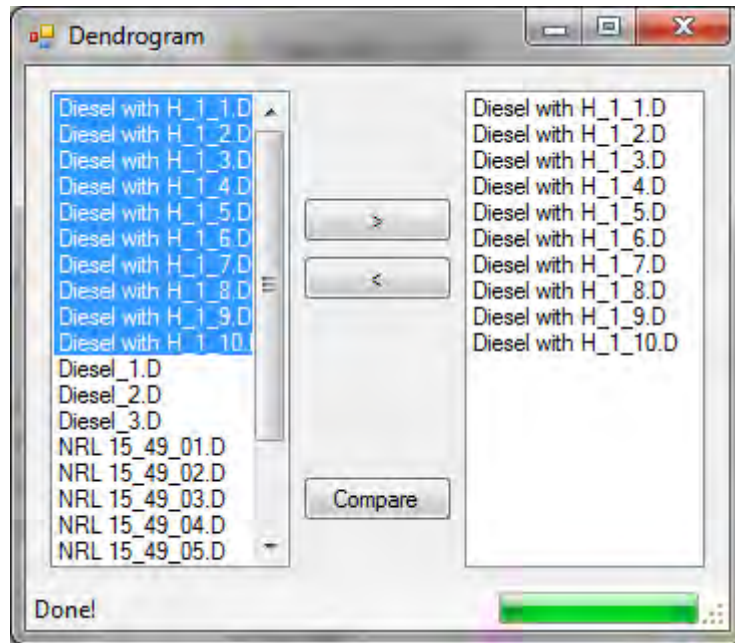
Figure 10. Dendrogram screen selection , showing the method for selecting the data to analyze (>) add to selected data (<) remove from selected data and (Compare) to begin cluster analysis.
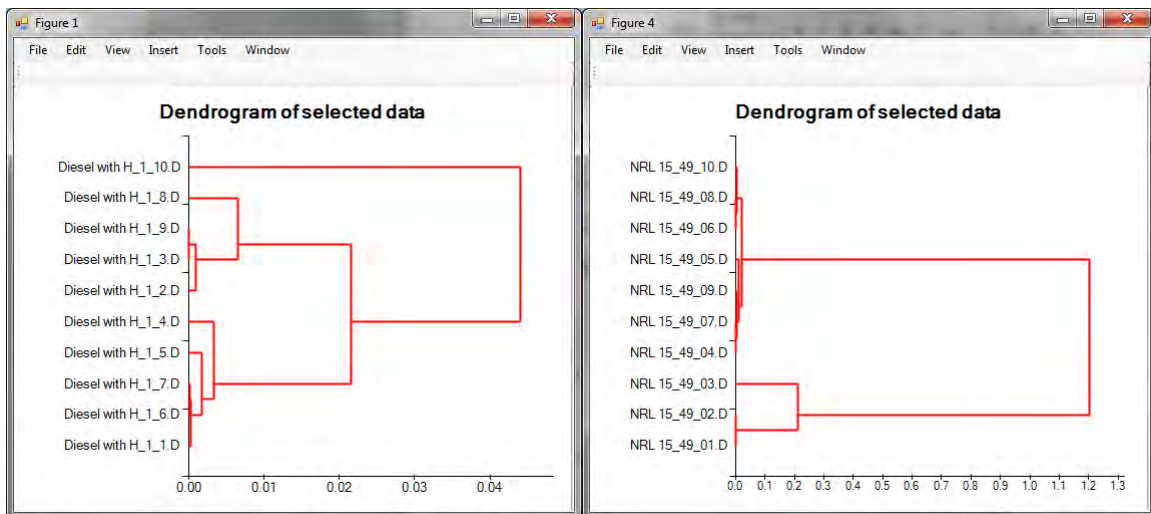

Figure 11. Dendrogram results screen, showing two examples. The results on the left show a strong similarity between all the samples with a cluster difference less than 0.1. The results on the right show a strong difference with three groups, consisting of 2, 1 and 7 samples, with a very strong difference between the first 3 samples and the remaining 7.

23

5.3 Settings

Search Parameters: These settings (figure 9) allow the user to select the search parameters for the compositional analysis. Selecting Reset, returns the settings to their default values. The file currently viewed will be reloaded to account for any change in the Min Match Factor in the display of the profiled hydrocarbon results.
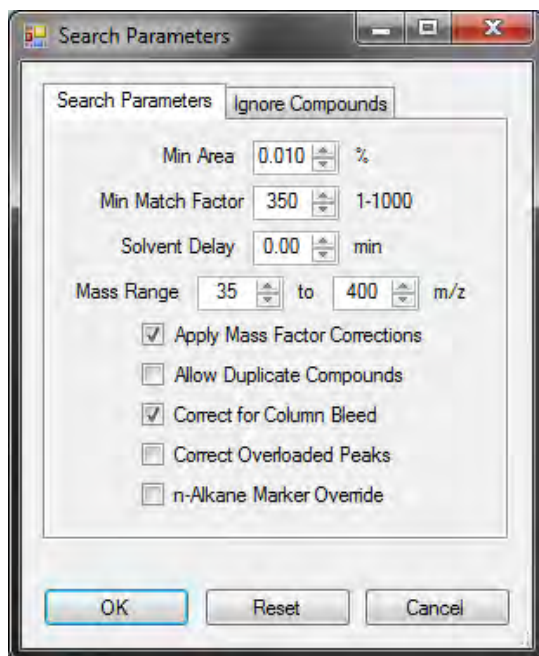

Figure 12. Dialog for setting compositional profiler peak search parameters.

- **MinArea:** The minimum area for a component to be added to the profile (default = 0.010%)

- **Min Match Factor:** The minimum match factor from the NIST MS Search for the component to be added to the profile (default = 850).

- **Solvent Delay:** The number of minutes to exclude from the data at the beginning of the acquisition to account for solvent elution (default = 0)

- **Mass Range:** The minimum/maximum mass range to use for the NIST MS Search, constrained by the mass range used to acquire the GC-MS data (default = 35 to 400).

- **Apply Mass Factor Corrections:** When processed data is loaded, peak areas are converted to mass percent with the appropriate compound class mass factors.

- **Allow Duplicate Compounds:** The reported profile listing will identify all peaks, and not combine peaks with the same name. This is only done when the data file is first processed.

- **Correct for Column Bleed:** The GCMS data file will be loaded, and the last 200 spectra will be used to do a baseline correction of the chromatogram to account for column bleed.

- **Correct Overloaded Peaks:** The GCMS data file will be loaded and any moderately overloaded peaks will be adjusted to correct for overloading.

- **n-alkane Marker Override**: This setting enables the user to ignore the retention times of the n-alkane compounds in the sample (if there are any) and use the saved list of n-alkane retention times to determine the distillation point profile.

Ignore Compounds: The ignore compounds tab allows the user to add specific compounds or name fragments that will be dropped from the profile if returned by the NIST MS Search. These include methylene chloride, siloxane, silane, silcoc, silyl and trifluoro as the initial default list. Care must be taken not to add any fragment (or name) to the list that is a valid compound that should be reported. For example "fluor" would be a poor choice to remove fluorine containing compounds, since flourene is a polycyclic aromatic hydrocarbon that would also be removed. Additionally the user may select the minimum number of m/z masses needed to be a valid compound, otherwise the program will skip the search for those below the set threshold.

Profile Order: Allows the user to specify whether the compounds listed under each compound class in the compositional profiles are sorted with respect to Abundance or Retention Time.

n-Alkane Marker Calibration: Allows the user to set the retention time of the n-alkanes (C6-C24) that are used in the simulated distillation calculations. The display (Figure 12) shows the TIC of the currently selected sample, indicating the retention times of the identified n-alkane compounds. The retention times are color coded to indicate whether they appear to be in the correct locations and are therefore used in the distillation calculations. Dark red lines are determined to be correctly located, while the light red lines appear to be incorrect and are ignored. The green lines throughout the sample are the saved calibration times that are used if the n-alkane Marker Override option is selected.

The saved calibration times should be adjusted to match the method used for any data where the override option is enabled. The average offset of retention times is shown below the displayed TIC, as well as the option to apply that delta to the saved calibration. To change individual Calibration retention times, just adjust the numbers in the table. Closing the window via **OK** will save the Calibration RT data, while clicking **CANCEL** will ignore any changes made.
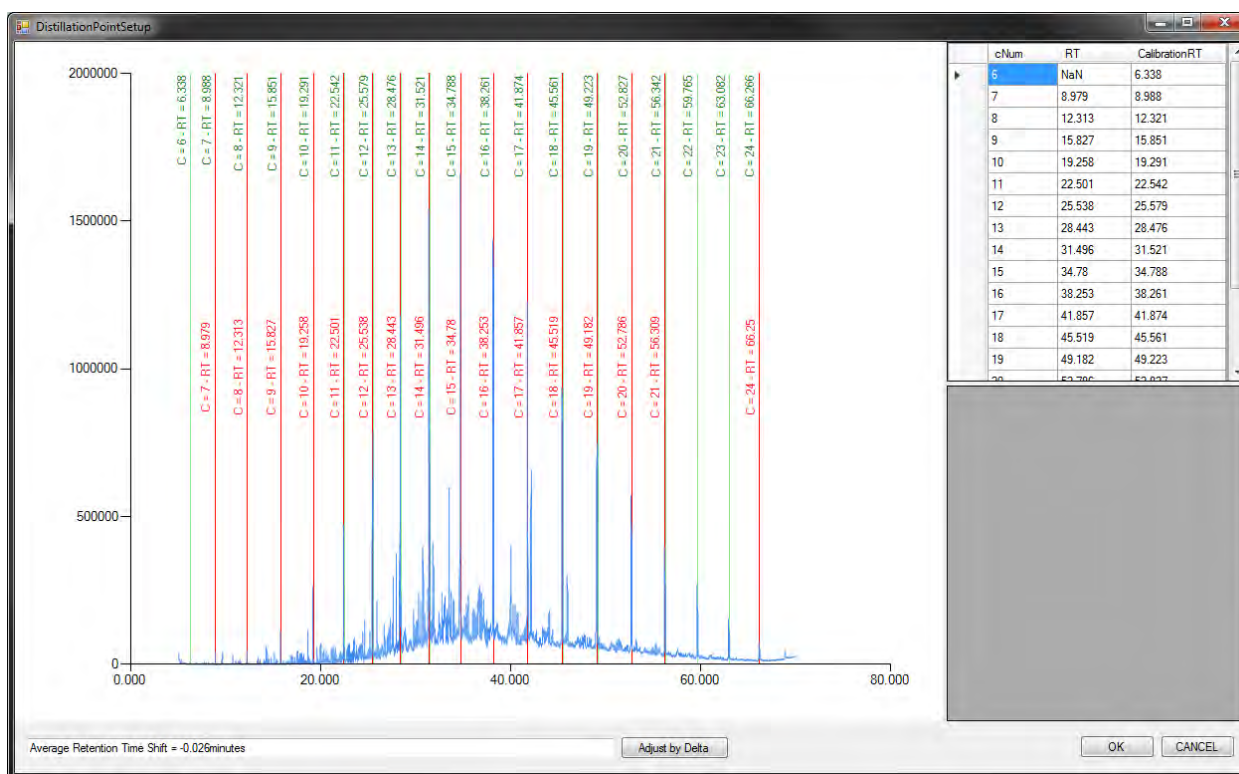
Figure 13. Interface screen showing n-alkane retention times in diesel fuel. The green lines show the saved calibration values available, whereas the red lines identify the retention times determined by the sample being analyzed.

5.4 Help

About: The about screen (Figure 13) shows the current version of the FCAST Software as well as the current version of the property models being used. These version numbers are saved into the results files as the data is processed, so a record is kept as to how and when the samples were processed.

26

Figure 14. FCAST information window, showing versions of the application and property models used.

ChangeLog: Describes the changes in the software since the previous versions.

6.0   Output of Processed Results

Exported Data. The results of processed data files can be exported to an Excel spreadsheet. The data exported consists of a summary sheet that contains for all exported files, the filename, sample name, noise factor, number of components found, a measure of data overloading, and which version of the software/property models was used. It also contains the profiler output, which consists of a summary of abundance in the major compound classes (saturates, aromatics, olefins, heteroatomics), compounds and their abundances in volume percent for each defined compound class, degrees of unsaturation (0-11), carbon number distributions (average, C_6-C_{28}) and the calculated properties.

Each individual sample that was exported also has a tab that contains more specific details about that sample. In addition to the information listed on the summary sheet. The report is broken down into several sections:
- general sample identification information
- area % by hydrocarbon class
- 20 largest peaks
- area% by degrees of unsaturation
- carbon profile by hydrocarbon class
- component listing by hydrocarbon class
- calculated properties
- plot of the total ion chromatogram

GCMS information Screen. The main screen of the FCAST (Figure 14) was designed to provide the analyst with an informative overview of the composition and properties of the processed fuel GC-MS data file. A variety of types of information are displayed about the selected sample, including the predicted properties, composition, total ion chromatogram, as well as the mass spectrum and mass fragmentation pattern of any selected fuel constituent. The slider in the Properties section allows the user to choose whether to evaluate the predicted properties against the relevant specifications for Jet or Diesel fuel. The property values are shown as green (in spec), red (out of spec), or black (no spec available). Any predicted property values that are not considered valid, are not displayed (NaN).

The list of data files indicates if a file has been processed by showing that entry as bold. The status bar at the bottom of the screen shows the data directory selected, the number of samples processed and the total number in the directory. The right side of the status bar contains a progress bar used in many aspects of the program.
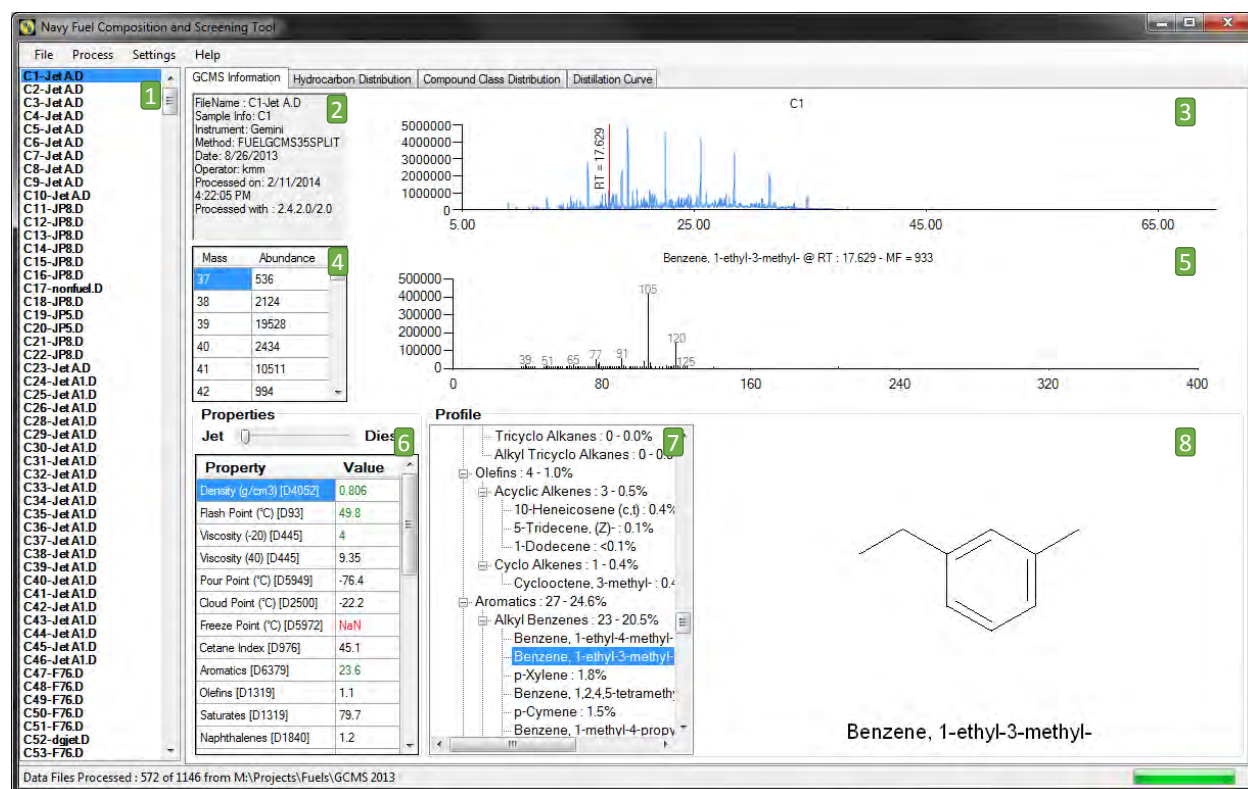

Figure 15. FCAST GCMS Information screen, showing 1) List of data files; 2) GCMS data file properties, as well as the date the file was processed with FCAST; 3) TIC of selected file, showing selected retention time of the selected compound; 4) m/z table for selected compound; 5) m/z plot for selected compound; 6) Calculated Properties of the sample; 7) Compositional profile in area percent; 8) Chemical structure of the selected compound in the hydrocarbon profile.

Hydrocarbon Distribution Screen. This screen (Figure 15) displays various information about the hydrocarbon distribution of the sample. The upper table shows area percentages for All CxHy/Saturates/Olefins/Aromatic by carbon number. By selecting the rows of this table, the bar graph changes to display the carbon number distribution of the saturates, olefins or aromatics detected in the fuel sample. Additionally, the degrees of unsaturation by carbon number in the fuel are shown.
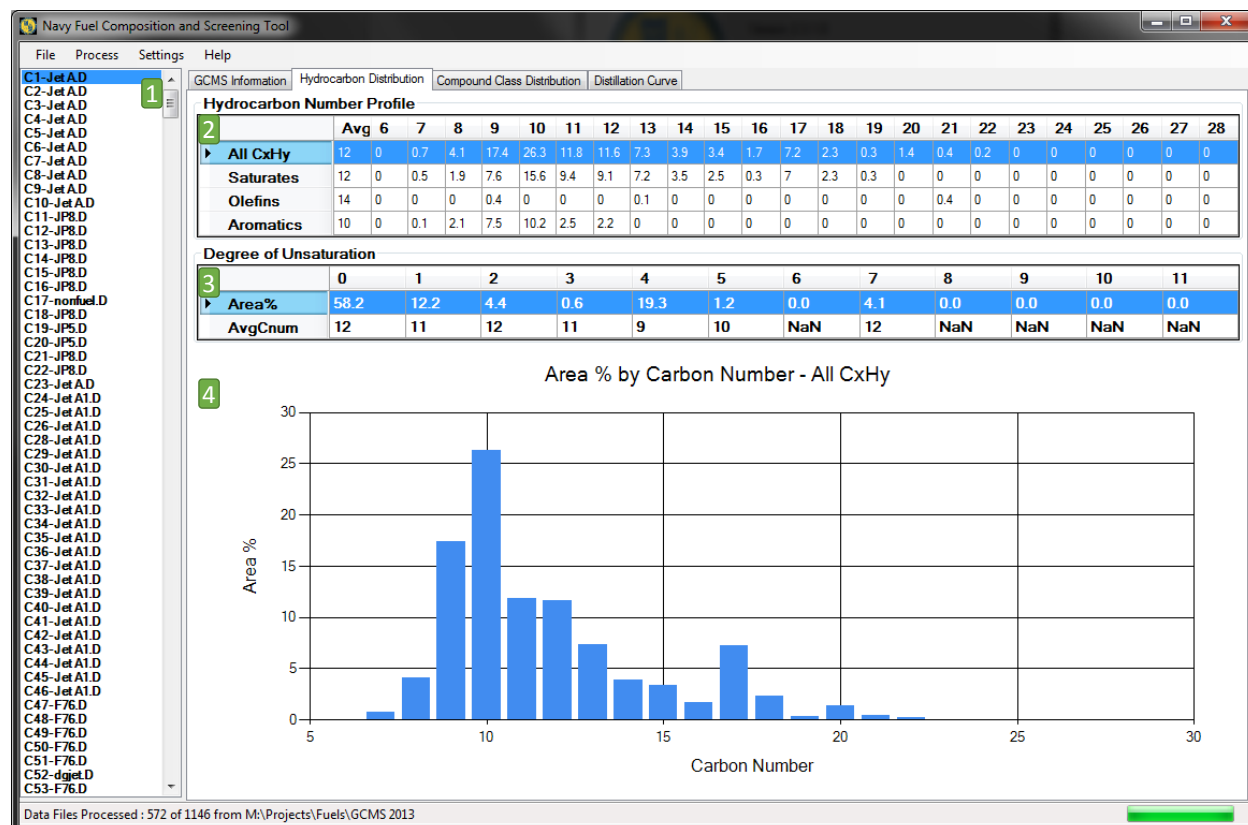

Figure 16. FCAST Hydrocarbon Distribution screen, showing 1) List of data files; 2) Carbon number distributions in area percentages for different classes of hydrocarbons in the sample; 3) Bar graph depicting the carbon number distributions in a selected compound class (selectable via the upper table).

Compound Class Distribution Screen. This screen (Figure 16) displays the abundance of different compound classes in a stacked bar graph. The operator can select which compound classes are displayed on the right side of the graph and the bar graph changes to display the carbon number distribution of each of the selected compound classes in the fuel sample. A context menu is available by right-clicking on the graph to choose all or none of the compounds as well as changing the colors of the bars displayed for each class. By right clicking on the graph, the operator can copy the plot to the clipboard for export to other applications.
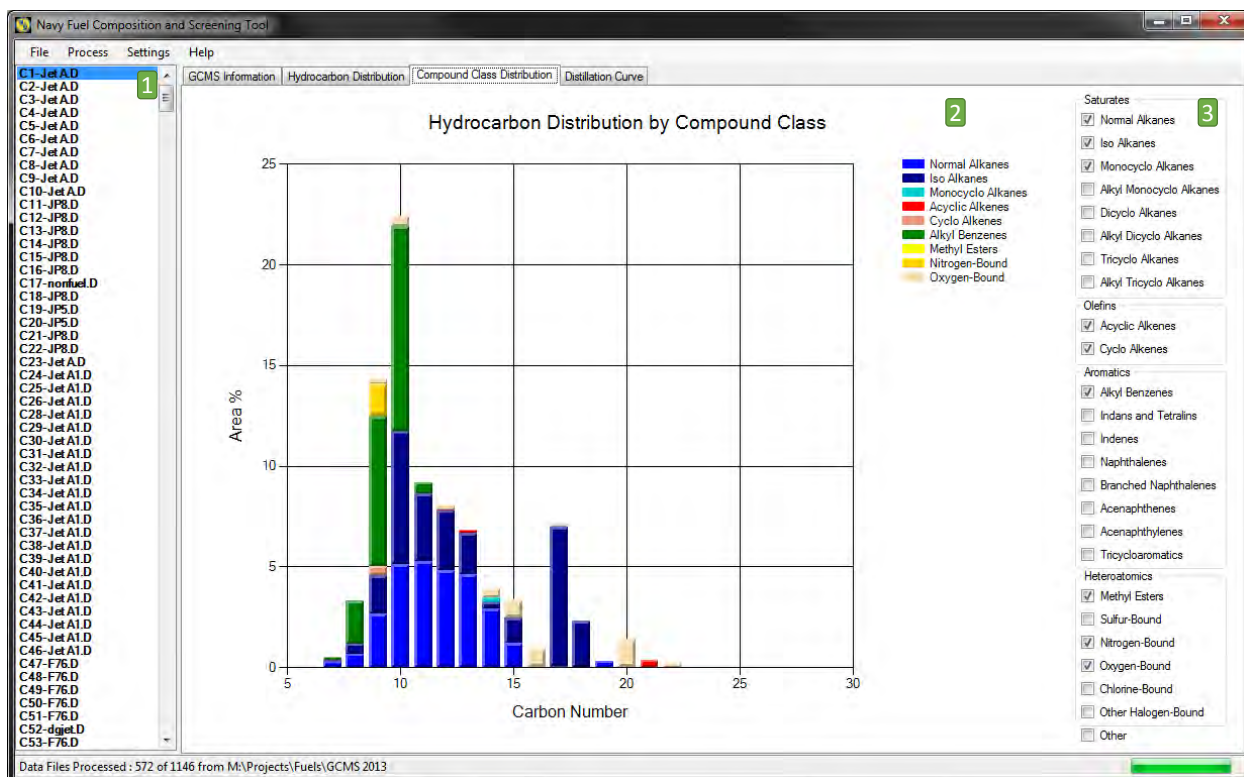

Figure 17. FCAST Hydrocarbon Distribution screen, showing 1) List of data files; 2) Stacked Bar graph depicting the carbon number distributions in the selected compound classes 3) Compound class list as check boxes to add or remove from bar chart.

Distillation Curve Screen. This screen (Figure 17) displays a predicted distillation curve using the same algorithm as the property calculations for the distillation points. The current sample selected is displayed with a black line along with typical jet and diesel distillation curves, for reference. If there is insufficient alkane peaks to calculate the temperature access, the screen will indicate that with an "insufficient data to generate plot" warning.
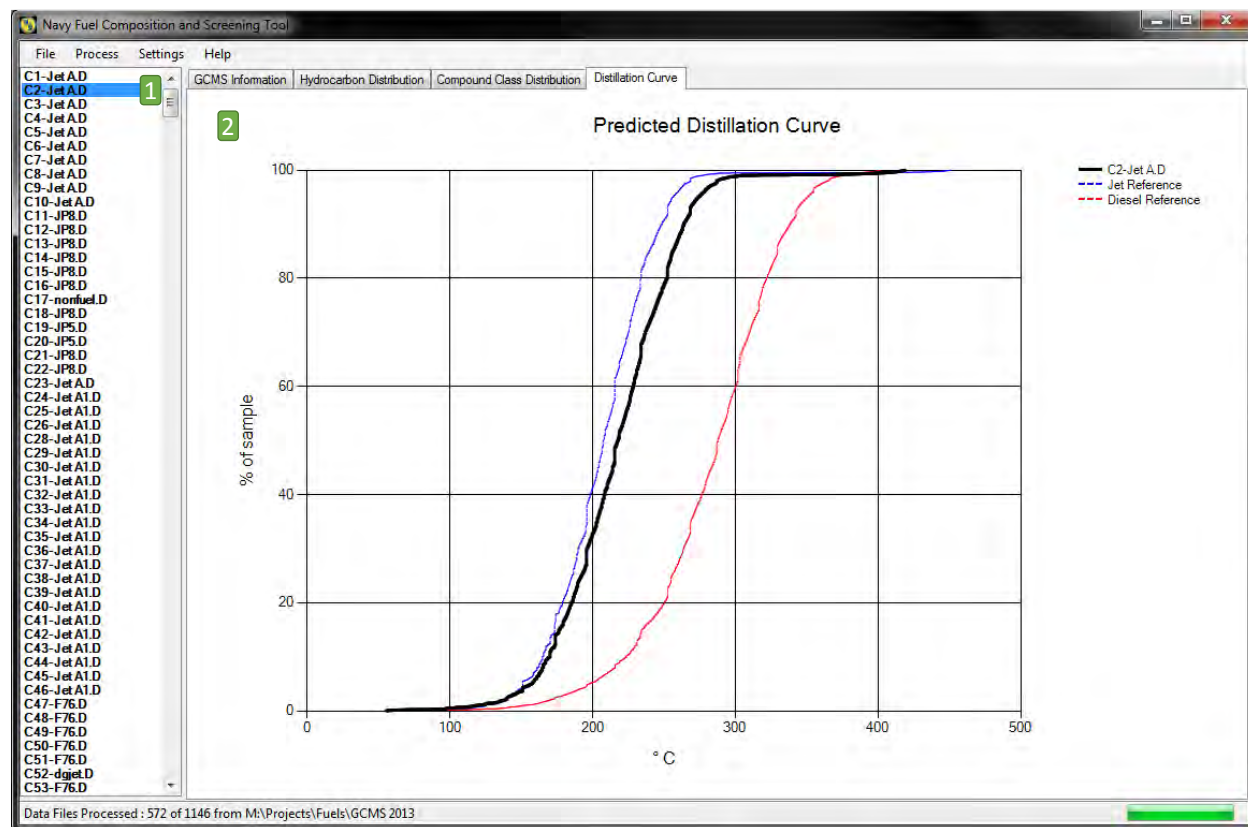

Figure 18. FCAST Distillation Curve screen, showing 1) List of data files; 2) Predicted distillation curve shown in black, along with jet and diesel reference curves.

Label Peaks Screen. This screen (Figure 18) displays the TIC with labels, showing the names of compounds with retention times. The selection tree on the right side allows for selecting/de-selecting labels based on major classes, minor classes, or even individual compounds.
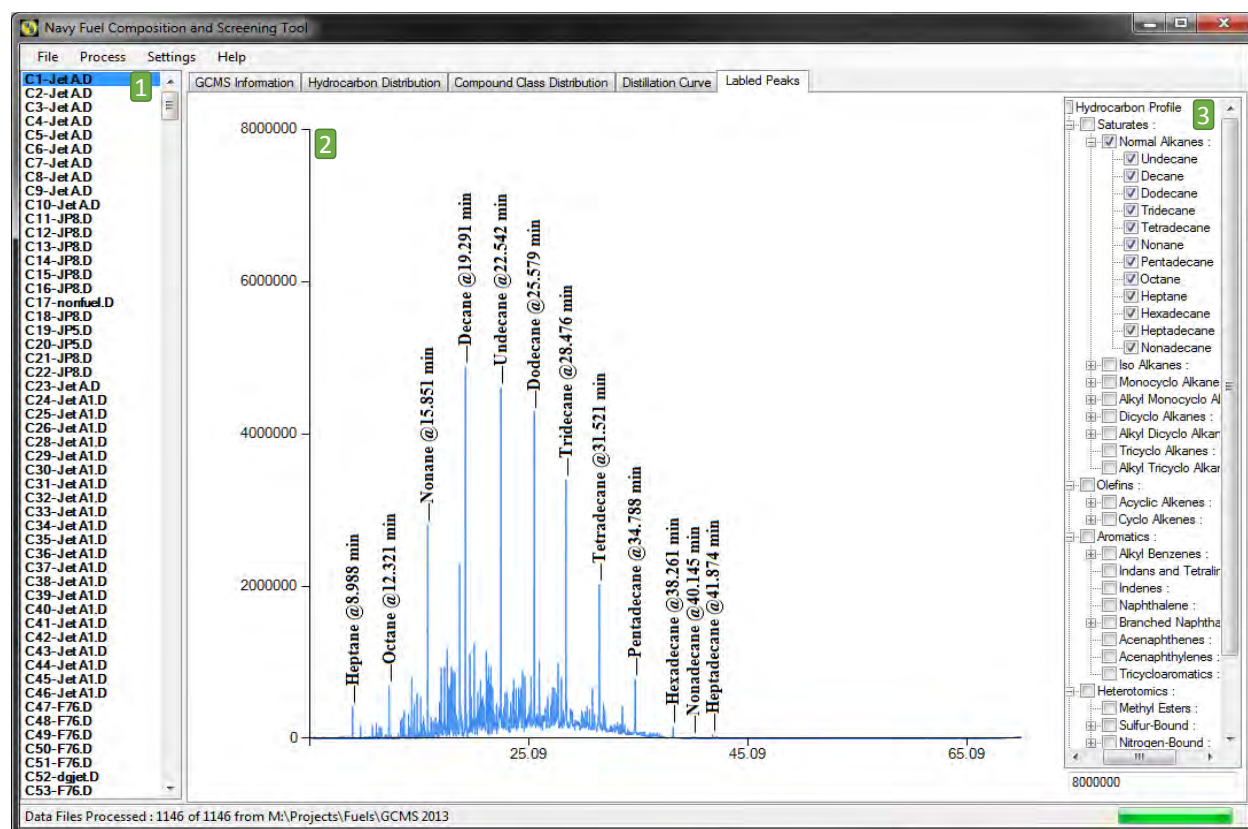

Figure 19. FCAST Label Peaks screen, showing 1) List of data files; 2) TIC with labels based on profile; 3) Selection tree enabling choices of either compound classed, or individual compounds.

7.0 Acknowledgements

8.0   Literature Cited

1. Hammond, M.H.; Morris, R.E.; Cramer, J.A.;Loegel, T.N.;Johnson, K.J.; Myers, K.M. "Navy fuel Composition and Screening Tool (FCAST) v.2.5. *NRL Memorandum Report No. NRL/MR/6180—14-9551*, July 18, 2014.

2. Begue, N.J.; Cramer, J.A.; Von Bargen, C.; Myers, K.M.; Johnson, K.J.; Morris, R.E. *Energy and Fuels* **2011**, *25*, 1617-1623.

3. ASTM Standard D2425. Standard Test Method for Hydrocarbon Types in Middle Distillates by Mass Spectrometry; ASTM International: West Conshohocken, PA, 2009; DOI: 10.1520/D2425-04R09, www.astm.org.

4. Standard Work Package, Naval fuels & Lubricants CFT Shipboard Aviation Fuel, JP-5, Qualification Protocol for Alternative Fuel / Fuel Sources", SWP44FL-006, Naval Fuels & Lubricants CFT, 04 March 2011.

5. Standard Work Package, Naval fuels & Lubricants CFT Shipboard Qualification Protocol for Alternative Fuel / Fuel Sources", SWP44FL-005, Naval Fuels & Lubricants CFT, 16 February 2011.

6. Morris, R. E.; Begue, N. J. "Compositional Comparison of Camelina Derived Jet Fuels to Their Petroleum Derived Counterparts". *NRL Ltr. Rpt.* 6180/0129, 07 June 2011.

7. Morris, R. E.; Begue, N. J. "Compositional Comparison of Algae Derived Jet Fuels to Their Petroleum Derived Counterparts". *NRL Ltr. Rpt.* 6180/0128, 07 June 2011.

8. Fitch, W., Sauter, A., *Anal. Chem.*, **1983**, *55*, 832-835.

9. Liu, G.; Wang, L.; Qu, H.; Shen, H.; Zhang, X.; Zhang. S.; Mi, Z. *Fuel* **2007** *86*, 2551.

10. Hupp, A. M.; Marshall, L. J.; Campbell, D. L.; Smith, R. W.; McGuffin, V. L. *Anal. Chim. Acta* **2008**, *606*, 159.

11. Fernandez-Varela, R.; Andrade, J. M.; Muniategui, S.; Prada, D.; Ramirez-Villalobos, *F. Water Res.* **2009**, *43*, 1015.

12. Sun, X; Zimmerman, C. M.; Jackson, G. P.; Bunker, C. E.; Harrington, P. B. *Talanta* **2011**, *83*, 1260.

13. Pedroso, M. P.; Fonseca de Godoy, L. A.; Ferreira, E. C.; Poppi, R. J.; Augusto, F. *J. Chromatogr. A* **2008**, *1201*, 176.

14. Zeng Z. D.; Hugel, H. M; Marriott, P. *J. Anal. Bioanal. Chem.* **2011** *401*, 2373.

15. Zorzetti, B. M.; Harynuk, J. *J. Anal. Bioanal. Chem.* **2011**, *401*, 2423.

16. Niu, Y.; Zhang, X.; Xiao, Z.; Song, S.; Eric, K.; Jia, C.; Yu, H.; Zhu, J. *J. Chromatogr. B* **2011** *879*, 2287.

17. Yang, L.; Bennett, R.; Strum, J.; Ellsworth, B. B.; Hamilton, D.; Tomlinson, M.; Wolf, R. W.; Housley, M.; Roberts, B. A.; Welsh, J.; Jackson, B. J.; Wood, S. G.; Banka, C. L.; Thulin, C. D.; Linford, M. R. *Anal. Bioanal. Chem*. **2009**, *393*, 643.

18. Aishima, T. *J. Chromatogr. A,* **2004**, *1054*, 39.

19. Jalali-Heravi, M.; Parastar, H.; Sereshti, H. *Anal. Chim. Acta* **2008**, *623*, 11.

20. Amador-Muñoz, O.; Villalobos-Pietrini, R.; Aragón-Piña, A.; Tran, T. C.; Morriso, P.; Marriott, P. J. *J. Chromatogr. A* **2008**, *1201*, 161.

21. Huang, X.; Shao, L.; Gong, Y.; Mao, Y.; Liu, C.; Qu, H.; Cheng, Y. *J. Chromatogr. B* **2008**, *870*, 178.

22. Marshall, L. J.; McIlroy, J. W.; McGuffin, V. L.; Smith, R. W. *Anal. Bioanal. Chem*. **2009**, *394*, 2049.

23. Song, S.; Zhang, X.; Hayat, K.; Jia, C.; Xia, S.; Zhong, F.; Xiao, Z.; Tian, H.; Niu, Y. Sens. *Actuators B,* **2010**, *147*, 660.

24. Miao, L.; Cai, W.; Shao, X. *Talanta* **2011**, *83*, 1247.

25. Bernabei, M.; Reda, R.; Galiero, R.; Bocchinfuso, G. *J. Chromatogr. A* **2003**, *985*, 197.

26. Kaspar, H.; Dettmer, K.; Gronwald, W.; Oefner, P. J. *J. Chromatogr. B* **2008**, *870*, 222.

27. Cramer, J. A.;Begue, N. J.; Morris, R. E. *J. Chromatogr. A,* **2011***, 1218,* 824-832

28. Booksh, K. S.; Kowalski, B. R. *Anal. Chem.* **1994**, *66*, 782A.

29. Beebe, K. R.; Pell, R. J.; Seasholtz, M. B. "Chemometrics: A Practical Guide"; Wiley, New York, NY, 1998, pp. 93-94.

30. Anderssen, E.; Dyrstad, K.; Westad, F.; Martens, H. *Chemom. Int. Lab. Sys.* **2006**, *84*, 69.

31. Gidskehaug, L.; Anderssen, E.; Alsberg, B. K. *Chemom. Int. Lab. Sys*. **2008**, *93*, 1.

32. Esbensen, K. H. Geladi, P. *J. Chemom.* **2010**, *24*, 168.

33. Haaland, D. M.; Thomas, E.V. *Anal. Chem.* **1988**, *60*, 1193.

34. Thomas, E. V. *J. Chemom.* **2003**, *17*, 653.

35. Lin, W. Q.; Jiang, J. H.; Shen, Q.; Shen, G. L.; Yu, R. Q. J. *Chem. Information & Modeling* **2005**, *45*, 486.

36. Centner, V.; Massart, D. L.; de Noord, O. E.; de Jong, S.; Vandeginste, B. M.; Sterna, C. *Anal. Chem.* **1996**, *68*, 3851.

37. ASTM Standard D86. Standard Test Method for Distillation of Petroleum Products at Atmospheric Pressure; ASTM International: West Conshocken, PA, 2012; DOI: 10.1520/D0086-12, www.astm.org.

38. Johnson, K. J.; Synovec, R. E. *Chemom. Intell. Lab. Sys.* **2002**, *60*, 225-237.